International Journal of Interactive Communication Systems and Technologies
Volume 9 • Issue 2 • July-December 2019

Categorizing Live Streaming Moderation Tools:
An Analysis of Twitch

Jie Cai, New Jersey Institute of Technology, Newark, USA
Donghee Yvette Wohn, New Jersey Institute of Technology, Newark, USA

ABSTRACT

Twitch is one of the largest live streaming platforms and is unique from other social media in that it supports synchronous interaction and enables users to engage in moderation of the content through varied technical tools, which include auto-moderation tools provided by Twitch, third-party applications, and home-brew apps. The authors interviewed 21 moderators on Twitch and categorized the current features of real-time moderation tools they are using into four functions (chat control, content control, viewer control, settings control) and explored some new features of tools that they wish to own (e.g., grouping chat by languages, pop out window to hold messages, chat slow down, a set of buttons with pre-written/pre-message content, viewer activity tracking, all in one). Design implications provide suggestions for chatbots and algorithm design and development.

KEYWORDS

Bot Design, Content Moderation, Human-Computer Interaction (HCI), Moderators, Online Community, Twitch Chatbot, Twitch Mod

INTRODUCTION

Live streaming is a mixed media form (Hamilton, Garretson, & Kerne, 2014) that is different from traditional social media in that it is considered a synchronous media with unique attributes such as simultaneity (Scheibe, Fietkiewicz, & Stock, 2016) and authenticity (Tang, Venolia, & Inkpen, 2016) and allows users (broadcasters and viewers) to interact with each other in real time through live video and chat (Wohn, Freeman, & McLaughlin, 2018). The live streaming platform, Twitch, is one of the leading live streaming video service providers that originally focused on games but is increasingly extending to creative content and mobile broadcasting. As of September 2018, Twitch had numerous content categories including IRL (in real life), Creative, Food & Drink, and Travel & Outdoors (Roger, 2018). The streamers are content creators and broadcasters of gameplay or other categories; viewers watch the streaming video then send messages to the streamer or other viewers in a chat interface that is adjacent to the streaming video.

DOI: 10.4018/IJICST.2019070103

The popularity of live streams and the success of Twitch have made it a growing subject of academic attention. Most current research on live streams, however, focuses on streamers and viewers, such as streamer or viewer motives (Cai & Wohn, 2019; Cai, Wohn, Mittal, & Sureshbabu, 2018; Friedländer, 2017; Scheibe et al., 2016) and streamer-viewer interactions (Lu, Xia, Heo, & Wigdor, 2018; Wohn et al., 2018), with less but growing attention on the prominent but hidden role of human moderators (Seering, Wang, Yoon, & Kaufman, 2019; Wohn, 2019).

Prior research defines content moderation as "the organized practice of screening user-generated content posted to internet sites, social media, and other online outlets, in order to determine the appropriateness of the content for a given site, locality, or jurisdiction" (Roberts, 2017). Generally, moderators perceive their roles as "filter, firefighter, discussion leader, and content expert" and they moderate content to guide the discussion and to keep down "flames" (Berge & Collins, 2000). Commercial content moderators, who are paid workers, curate content and guard against violations such as racism, homophobic slurs, pornography, and violence (Roberts, 2016). These commercial content moderators usually review inappropriate content that has been flagged by users or detection algorithms. Twitch has commercial content moderators but also enables streamers to appoint their own moderators (also known as "mods"). Mods voluntarily assist the streamer in managing the chat content and are usually unpaid (Wohn, 2019).

Due to the synchronicity of live streams, all the messages are flowing in the chatroom in real time, posing different challenges compared to asynchronous communities such as Wikipedia and Reddit, which also rely largely on volunteer moderators. Technical interventions can, to some extent, reduce the human moderation load, especially in large and fast-moving chats (AnyKey, 2016). Many online communities, such as Reddit and Twitch, apply bots (software robots) to assist the mods in doing moderation practice (Seering, Wang et al., 2019). Current research about using bots for content moderation mainly focus on asynchronous communities such as Reddit (Gilbert, 2013; Long et al., 2017) and Wikipedia (Clément & Guitton, 2015; Müller-Birn, Dobusch, & Herbsleb, 2013), with limited research about bots for moderation on Twitch (Seering, Luria, Kaufman, & Hammer, 2019). Better understanding the moderation tools that mods use every day would help improve the current tool design, reduce the working load of mods, and further benefit the community. The goal of this research is to analyze the features of moderation tools on Twitch into categories that could be generalizable to all other moderation tools and to provide some implications for future tool design.

BACKGROUND

Online Community Moderation and Moderators

One of the primary reasons that online spaces need moderation is because of the negativity that persists regardless of platform. Racism, sexism, and many other prejudices flourish online (Chadwick, 2006) while trolling, flaming, spamming, and flooding messages can disrupt users' experience (Lampe, Zube, Lee, Park, & Johnston, 2014; Pfaffenberger, 2011). Seven key risk categories need to be addressed by moderation techniques for user-generated content: offensive content, spam, soft hacking, etiquette breach, editorial conflict, copyrighted material, and personal exposure (Coutinho & Jose, 2017).

Communities can use formal and informal methods to enforce standards of appropriate behaviors such as explicit rules, reputation systems, and algorithms (Anis, Börrnert, Rümmeli, & Kuntscher, 2013). Some research suggests that new IT technologies such as moderation systems could deal with information overload and improve participants' civility (Mironov, Faizliev, Sidorov, & Gudkov, 2016; Resnick, 2002). Researchers have investigated tools to mitigate information overload and manage negativity. Crowdsourcing uses the ability of a large group of people to provide feedback about a single piece of information—for instance, by flagging inappropriate content that is then overviewed by human moderators (Resnick, 2002). Crowdsourcing is commonly employed through rating systems and

used for many online communities such as Facebook and Twitter (Gillespie, 2018), Amazon (Gilbert & Karahalios, 2010) and Slashdot (Lampe & Resnick, 2004; Lampe et al., 2014). Crowdsourced methods, however, are sometimes inaccurate or untrustworthy (Ghosh, Kale, & Mcafee, 2011) and run the danger of not being fast enough because there is often a time gap between when someone reports "bad" content and when the moderators review it.

Algorithms can help with the detection of troublesome content, but they are not perfect (Seering, Kraut, & Dabbish, 2017) and often need to be overseen ultimately by a human moderator (Wohn, 2019). Bots are part of these algorithms, but the quality and functionality of bots still pose some social and practical challenges (Long et al., 2017). On Wikipedia, a bot that proactively enforced the guidelines and norms caused polarized (either positive or negative) responses from users (Clément & Guitton, 2015). Due to a large amount of content, however, no single approach is effective, and a combination of both algorithms and labor is the current approach for many platforms (Lampe & Resnick, 2004; Roberts, 2016).

Social Media Moderation

Moderation on social media can have elements of censorship (King, Pan, & Roberts, 2013) but can also facilitate or encourage certain types of behavior. Research of moderation on Instagram about pro-eating disorder (pro-ED) found that non-standard lexical variation of moderated tags has emerged over time; these variant tags even expressed more toxic, self-harm, and vulnerable content; and the participation and support of pro-ED thrived and increased (Chancellor, Pater, Clear, Gilbert, & De Choudhury, 2016). Daniel et al. (Daniel, Bernd, & Tom, 2013) designed a workflow for Twitter to integrate disaster management system and employed content moderation to ensure the quality of the disseminated information. Content moderation for social media platforms and commercial sites ensured brand protection, adherence to terms of use statements, and site guidelines and legal regimes (Roberts, 2014).

The moderation techniques on social media could be categorized as pre-moderation, post-moderation, automated moderation, and distributed moderation; scholars have suggested that different types of user-generated content should employ different types of moderation (Veglis, 2014). For example, comments should use distributed moderation, forums should use pre-moderation, but social media is a complex issue and hybrid moderation, which is a mix of all moderation types, should be employed. Coutinho and Jose (Coutinho & Jose, 2017) categorized moderation approaches by entity difference: the display owner, the system itself, a set of trusted curators, and accountable publishers; seven moderation approaches were classified: content pre-approval, automated filters, delegated content curation, social accountability, content removal, distributed content removal, and report abusive content.

Summarizing the research on moderation in social media and online communities, the authors found that there has been much discussion about the labor aspect of moderation, the automation of moderation (Delort, Arunasalam, & Paris, 2011; Hammer, 2017; Saúde, De Medeiros Soares, Basoni, Ciarelli, & Oliveira, 2014), and the different approaches to moderation. Recent research about moderation in live streams focused on the motivation being a mod (Wohn, 2019) and how mods engage with their communities (Seering, Wang, et al., 2019). However, there has been relatively less discussion about the individual practice of human moderation concerning the different tools that are used by the moderators. This less discussion may be in part because many social media platforms do not offer moderators much autonomy aside from deleting malicious content. In that sense, the context of live streaming on Twitch is particularly interesting from a moderation perspective because moderators can have more direct engagement with different technologies that enable them to perform a range of functions. Thus, the authors asked:

RQ1: What kind of moderation tools do Twitch mods use in live streams?

Twitch employs a multi-faceted approach to moderation. The company itself employs human moderators who mostly handle moderation of content that has been reported by users as being inappropriate. It also has a moderation tool called AutoMod, a tool that uses algorithms to help streamers moderate their chatrooms. AutoMod performs various functions, especially at a preventative level, such as preventing people from typing in certain offensive words, preventing people from posting links, or preventing spam. This proactive moderation tool on Twitch could effectively discourage spam and specific types of negative behaviors (Seering et al., 2017), but it fails to fulfill all the moderation needs of moderators. Thus, many moderators have to use a lot of other plugins or extensions to facilitate the moderation process. Twitch is a unique platform in this respect in that it allows users to utilize third-party tools to facilitate content moderation—this differs greatly from other types of social media and live streaming platforms where the company handles moderation centrally. The fact that many mods have to implement various moderation tools might indicate that the existing tools are incomplete and that the mods might need more functions of moderation tools. Thus, the authors posed the following research question:

RQ2: What do mods expect from moderation tools in the future?

METHOD

Participant Recruitment

Since the research questions are about how moderators use tools to moderate chat content, the authors targeted our participants to volunteer moderators on Twitch and used semi-structured interviews to ask them about their moderation experiences. A few different methods were used to reach out to them. The first way was through Twitter. The authors used the official Twitter account of their research lab to post recruitment messages, to search for profiles using keywords such as "Twitch, mod, and moderator," and to reach out to moderators by sending direct messages. Second, private Twitch accounts were used to reach moderators by directly messaging active moderators in random channels through Whisper (a message feature of Twitch). The authors also recruited moderators through streamers that were interviewed for a separate project. 21 Twitch moderators were recruited, and each moderator received a $20 gift card for their participation.

The interview protocol was reviewed and approved by IRB first; the semi-structured interview was 40-60 minutes through phone call or Discord (a communication application often used by streamers and moderators). The protocol included questions about general motivations such as why they mod, whom they mod for, and the different tools and methods they use for moderation. The authors also asked them at the end of the interview if there were any moderation features that they wished for the future. The interviews were audio-recorded for further data analysis with participants' permission. Audio transcriptions were completed and double-checked for accuracy by six research assistants and the authors.

The grounded theory consists of a set of inductive strategies for researchers to develop abstract conceptual categories to understand the qualitative data and identify the patterns within it (Charmaz & Belgrave, 2015). The summative content analysis started with identifying keywords or content to further interpret the context (Hsieh & Shannon, 2005). The authors combined these two approaches. The first author coded participants' descriptions relevant to the first research question as key concepts one by one, such as "ban and timeout" and "filter words." The second author reviewed the codes and discussed with the first author to clarify inaccuracy and ambiguity of some codes and to ensure consistency of the coding criteria. Then, the first author coded all the following interview questions following the established criteria and put all relevant quotes under each code for further review and discussion with other authors. Finally, all authors sat together and read each quote. Based on the

similarity of quotes and concepts, the authors grouped codes into different categories. After several rounds of grouping, the authors finally identified higher level themes.

Participant Demographics

Table 1 lists the main demographic characteristics of our participants. These characteristics indicate a diverse sample in this study. Results showed that most participants were male (71.5%), followed by the female (19%) and transgender (9.5%). The average age was 29, ranging from 18 to 45. The average moderation experience was two and a half years, ranging from one to five years. The number of channels they moderated was a wide range from one to eighty. Most mods moderated less than five channels (71%), one participant was very active and had a channel list that contained 80 channels. On average, they moderated 23 hours a week, varying from 2 to 84 hours.

RESULTS

Moderation Tools

Based on moderation tools that they used, moderators could generally be divided into heavy technology users or light technology users. Most of them were heavy users, and if they used bots, they usually

Table 1. Moderators' demographics and activities

	# of Channels That They Mod For	# of Years as a Mod	Age	Gender	# of Hours Spent per Week Moderating
P1	2	2-2.5	23	Male	21 - 84
P2	1	2	N/A	Trans	6
P3	6 or 7	5	31	Male	10
P4	80	4	24	Male	20
P5	30	3	21	Male	N/A
P6	2	N/A	43	Male	Depends
P7	2	1	33	Female	20
P8	1	2	18	Male	60-70
P9	A couple	N/A	N/A	Male	35-42
P10	1	1.5	37	Female	3
P11	2	1	20	Male	21-28
P12	1	1	21	Male	N/A
P13	60	2.5	41	Male	21-28
P14	2 or 3	1	29	Male	12-16
P15	44	2	19	Male	2-3
P16	20	2	40	Female	12
P17	4	3-4	40	Male	4-12
P18	3	4	N/A	Male	8-10
P19	5	5	27	Female	36-70
P20	1	1	45	Trans	16-24
P21	4	2	35	Male	30

used more than one and the combination varied. Some were light users and stated that they did not like bots and that the bots often caused more trouble so that they mainly moderated manually and only used the basic bot embedded in the system.

The most popular bots or extensions that our participants used were: Nightbot (38%), Twitch AutoMod (33%), Better Twitch TV (BTTV) (33%), Moobot (19%), individually developed bot (19%), and FrankerFaceZ (FFZ) (10%). Among these, only the Twitch AutoMod was built into the Twitch system; others were third-party plugins or extensions. (Although AutoMod is in the Twitch system, users can choose not to activate it if they do not want to use it). Interestingly, some participants mentioned they were using tools that they or their friends developed. Then, the authors categorized these tools regarding their features. Based on participants' description, four categories and nine examples of the features that fall into those categories are summarized in Table 2.

Chat Control

Some moderation features were associated with control of chat, a place where viewers could comment on streamers and communicate with each other. The chat interface is side by side to the live stream (on PC it is on the right, on mobile devices the chat is on the right or beneath the video, depending on whether the device is held vertically or horizontally) and happens simultaneously.

Because of the live interaction on Twitch, all the new messages sent by viewers would be automatically displayed at the bottom of the chat, making it challenging to go back and check chat history if new messages were constantly appearing. The inconvenience of going back caused difficulty for some mods. "When you go on Twitch, and you try to delete a message, and you scroll up, if somebody sends a new message it automatically goes to the new message," P1 explained. Some extensions could help them control the speed of the chat movement. P1 added: There is a tool that makes it when you scroll up it does not go back down." In big channels with lots of viewers, the chat moved so quick that they could not catch negative comments—for situations like this there was a feature that could make the chatroom still. P9 said:

I have an extension where if I hover over the chat with my mouse, it just stops the chat, so I can properly click on someone's name and moderate.

Content Control

Flagging and alerting "bad" messages was a feature mainly integrated into Twitch AutoMod.P5 explained this feature:

I… turn on AutoMod, which is Twitch's automation thing because all that does is flag messages as pending. So, if a message is deemed inappropriate by your channel, it'll flag it and then put it in chat for the moderators. They can say approve or deny.

Table 2. Moderation tool categories and examples of tool functions

Chat Control	Viewer Control
• Chat movement control: P1, P9 • Multi moderating: P18	• One click and purge: P1 • Ban or timeout: P2, P13 • Pause without timeout: P8 • Log view: P5, P18
Content Control	**Settings Control**
• Flag and alert message: P1, P5, P20 • Filter words: P2, P6, P18	• Customization: P2, P5, P8, P18

In addition to flagging and alerting messages, the system could also automatically filter certain words. Moderators or streamers could set and put filter words in bots so that these words or the variants of these words typed by viewers could not be displayed in the chatroom. "You can put specific words into it that just don't go through," said P2. P18 expressed his appreciation for this feature:

By far my favorite feature of AutoMod is whenever people send a message, it automatically doesn't go to the chat. What I really enjoy about automod is that it pretends [the message] doesn't exist, it turns it into a none and done a deal where no one saw it; no one is reacting; there's no drama- it's gone.

Viewer Control

There were many features in controlling viewers' behaviors. "One click and purge" allowed moderators to easily and conveniently delete the offensive message and "time out" viewers from the chatroom simultaneously. P1 said: "It is easier to purge people because it is just one click and you purge them or ban them whereas on Twitch you would have to actually like type it out with like purge or timeout or ban. So, it allows you to do things more conveniently." The ability to do something with "one-click" indicated the efficiency of using the moderation tool.

If someone said something inappropriate, some words that have been considered too toxic or offensive by streamers or moderators, the ban or timeout rule would apply. This feature was mainly implemented through extensions. "I use BTTV, and that gives some nice things to make it easier to time out and ban," P13 said. Nightbot also had a similar function, filtering words first and then timing out the person. P2 said:

Nightbot tries to make sure if someone says "faXXot" it just does not appear on Twitch. It just… that person will end up timed out. It automatically times out the person from being able to talk for a specific number of seconds. I believe it's 60; I'm not sure.

Pause without timeout was a little different from and less severe than a ban or timeout. Instead of timing out a person for a specific period, a pause would slow down the speed of messages that one could send. P8 said:

Instead of choosing to permanently ban somebody or time them out for 10 mins in chat, much time you will see a mod purge somebody, which is just literally to time them out for one second, and I have this setup … in my settings that I have a button to set people's name that I can automatically purge them without actually time out like slash timeout.

A pause without timeout worked as a light warning. The messages had no problem, but someone might want to get attention and, instead of typing a sentence that might be overwhelmed by others' messages, might type quickly word by word to take up multiple lines. Then the whole chatroom would be occupied by the messages. These messages would annoy other viewers and dilute community experience.

Log view allowed the moderators to check a specific viewer's log. By doing so, they could see the chat history of the viewer. "His most useful tool by far is what he calls a log viewer, which pretty much lets me pull logs from anytime a user has talked in a channel as long as it's been logged," said P18. Especially when some viewers were discussing lightly harmful topics, but the moderators had difficulty in deciding whether to give a warning, a timeout, or a ban. Checking logs would help moderators to make better decisions. P5 explained:

You can look up people, see how long they've been following. We can see previous chat messages; you can see all tons of information about them. So, whenever I see a new name in chat, I'll click them

and see how long they've been on Twitch. If it's a day one account, I'm immediately skeptic and I watch them like a hawk. Otherwise, I just let them chat.

Settings Control

Many moderators discussed that customization of settings based on their needs made moderation more efficient. "It is more efficient. You can customize the tool whichever way you want, and it's just a lot better for people," said P8. Some bots provided the option to customize timeout, for example. "A common plugin for Twitch, you can add custom timeout buttons for different tasks," said P5. Similarly, P8 said, "For external tools sometimes I use custom IRC clients if I want to run like a custom bot to look for a specific keyword to time out." Some bots allowed customized settings to track details of chat activities. P2 said:

When I created my own (setting), it's like, it's very detailed. It tells you everything that happened, even while you're not in the chat. Something that will happen a week ago, it'll be like this is what went down.

Even though current bots provided a certain level of customization, from our interviews, some moderators were not very satisfied with the performance of customization. More options for current features such as timeout settings could be considered higher-level customization as well. "The Twitch tool, it is mostly being able to do it one second, 10 seconds, or say one second, one hour, or 10 hours or whatever. That's pretty much it. Like it does need to be more in-depth than that," said P19. These deeply customized features would meet moderators' diverse needs, reduce their workload, and accelerate the moderation process.

Through the analysis of current moderation tools, nine features were highlighted, and four categories were identified. However, are these all they wanted? Are there any other features they expected? The following research question asked about moderators' needs.

The Desired New Features

Our second research question was about what mods desired in the future. The question specifically asked the mods in the interview, "If someone could design a moderation tool or bot for you, what would you want it to do?" Since not all moderators have used all existing tools in the market, some wanted features that already exist and were covered in the previous section. Thus, in this section, only new features not mentioned above will be discussed. Six features were identified: grouping chat by languages, having a pop-out window to hold messages, chat speed control, a set of buttons with pre-written/ pre-messaged content, viewer activity tracking, and all-in-one. Ironically, some of these features were already available with existing bots or extensions, but the participants were unfamiliar with it.

Grouping Chat by Languages

This feature was relevant to the content control category but different from any features mentioned above. There were many viewers from different countries, speaking different languages, but watching the same streaming event. Not all viewers would type and communicate in a single language. However, if the moderators only understood one language, it would be difficult for them to moderate when the content of different language mixed. Moderators might be distracted and have to pick out messages that they could read and understand, even though different moderators were assigned to handle different languages. Therefore, they wished to have a function to group different languages for different moderators. Doing so would improve moderation efficiency. Moderators also wanted translation abilities to help out with chat in different languages. P1 said:

Like, because Gears of War is so big in Mexico, and it's just a lot of people who speak Spanish are in the chat. Sometimes it gets overwhelming to the point where the American, or the people who speak English only. They might not have anything to do in the chat because we just can't understand what's being said. So maybe a feature on Twitch or Mixer that automatically [translates] Spanish, or any language in general, to English would be cool and helpful for us so that the people who only speak English, or not only speak English but predominantly speak English, could help along. It also helps the moderators who speak Spanish because now they have so much more work to do because it's not equally divided among us. So, they have a heavier workload.

Pop Out a Window to Hold Messages

This feature could be under chat movement control category but was different from the features mentioned above. A pop out window would hold the message that the moderator wanted but would not change the chat flow. The participant said Twitch once had this feature, but after the update to the latest version, it was gone. Now it was hard to hold messages. P7 said:

I think popout would be very good. If you could make it, so a bot could make a pop-out window so that when you click on something it would hold it. Now you don't get to pop out where you can inspect what the person is saying. If I could get a bot to bring that sort of thing back. Because if you can go back and look over the sort of things somebody saying if they're just swearing and it's a one-off assessing something inappropriate, it's a one-off. It's not such an issue, but if I can go back and see that this person has insulted X, Y, and Z, I think I said something inappropriate to someone and its little things, then you know, you've got to keep mind.

Chat Speed Control

This feature was also relevant to chat control but different from other features mentioned earlier. P8 said that the messages moved so quickly and were hard to catch up. However, he only hoped a new feature to slow down the speed so that he could not click and moderate by mistake. Something might look like an audio player, and there are options such as slow down, keeping normal, speed up. He explained:

The chat moves quickly, so you want to slow it down… If I want to timeout someone and someone posts, the chat is going to go up like one line, so I can ban someone else by mistake.

A Set of Buttons With Pre-Written/Pre-Message Content

This feature could be under settings control but was different from the customization features mentioned above. Again, some bots already have this feature, but the participants were unaware of them. Participants mentioned that they would need commands with pre-written information so that they could reply more quickly than just typing the same message again and again. "I would just press a button, and it would instantly reply with something, that I had pre-messaged or pre-written," said P11. With this setting, moderators could work more efficiently. P15 described his expectation and said:

I would probably have it be like go all around so it would probably have stuff I'd take inspiration from night bot you know having commands with info, so having that ready… obviously, it'll be quicker than us since it is a bot and not a person.

Viewer Activity Tracking

This feature could be under the viewer control category but is different from the log view. In the log view, moderators wanted to check one specific viewer's chat history and make better judgments based on the viewer's current performance. Viewer activity tracking was about the general behavioral summary of a group of viewers. For example, what percentage of them are super active? How many of them are lurking? How many new viewers joined in or left last week? "I would want something that would track everyone else. I want some vocal data about regular people, get notices if people do not show up. I can notice if people suddenly people get depressed, maybe that," said P21. Many moderators expressed their care about their viewers during the interview and considered some of the viewers as friends and had a good relationship with viewers. By owning this feature, moderators and streamers could have a better understanding of viewers' activities. Therefore, they could improve their service and maintain a better relationship with viewers and doing so would be beneficial to the community as well.

All in One

This was not a novel idea, but moderators wanted something that integrates all the features of moderation tools in the current market into one. P4 moderated for several big channels and had to use five bots to assist the moderation process because currently, no one tool could meet his requirements. He explained:

I think it'd be cool to have an all in one moderation bot where you can type in a name and give it like a Twitch whisper or something else, so you could pull it quicker than you could from going through a website or chat logs in a program.

DISCUSSION

The first research question identified four different perspectives taking the synchronous nature of live streaming into consideration, preliminary providing a guideline for further bot development in this domain, and the second research question supplements the four categories identified in the previous one. Similar to Seering et al.'s findings, viewers control involves a certain level of multiparty interaction between moderators and viewers. Future design can explore how to facilitate the interaction at the same time improve moderation efficiency. Our results also show that the moderation tools in synchronous online communities are different from these in asynchronous online communities such as Wikipedia and Reddit. For example, chat control involves real-time content management, and mods have to deal with information overload and to make decisions immediately, suggesting that mods in live streaming communities are undertaking a different type of time-sensitive psychological pressure than those in other communities.

The categorization of moderation tools enables us to think about features in a more systematic fashion, not only in identifying the different types of problems that exist, but also where more work needs to be done. According to the analysis of features of current moderation tools and features that moderators expected, the authors have several suggestions for the design of the platform as well as suggestions of new features.

Design Opportunities

Specifically, for Twitch, the leading live streaming platform, the main features of its AutoMod are mostly under the content control category, which means that features under the other three categories are opportunities for future development. Twitch allows third-party extensions, thus opening up opportunities for a myriad of different moderation tools. However, it is still difficult for beginners to choose which tools to use. The beginners might have to add so many extensions to test one by one

and then only keep the better ones. If Twitch can add a function to categorize tools by their features, it would be helpful for moderators, especially beginners, to search for the tools that they need.

Some moderators expressed the desire for features that already exist, indicating that searching for these third-party tools are inefficient or that there is a lack of information about where to find extensions or bots that are less well known. Future research may want to look into how moderators discover these tools, but the fact that people do not know about tools that already exists means there are more opportunities for centralized repositories of these tools and education about how to use them.

Technology updates so quickly. Some unavailable features during the research time are now available on Twitch. For example, some interviewees mentioned that when they scrolled up the chat, it would automatically go back down. However, now when scrolled up, the messages will stay where they are stopped. Twitch also has a "Popout" window to hold chat and to run separately. Moderators can keep both the chatroom embedded in a streaming webpage and the "Popout" window open and can use the chatroom to track general behaviors of viewers and the "Popouts" to deal with suspicious viewers. The evidence further exemplifies the importance of understanding the function of these features from a higher perspective than the feature themselves. The identified categories are not time-sensitive.

Suggested New Functions

Based on some of the frustrations and problems that moderators discussed, the authors suggest a couple of ideas for new features that could be applied to any live streaming platform.

Highlighting the content moderators want to track: a language setting button that allows moderators to choose what kind of language would be highlighted on their screen that will help them focus on what they can handle and increase working efficiency. For example, a Chinese moderator would only want to moderate Chinese content in the chat and click the button to show Chinese messages only. All the Chinese would be highlighted, and other language content would turn gray or shadowed so that they could concentrate on the moderation of Chinese content.

Instead of checking viewer's log (which would mean that during that time the moderator would be ignoring the whole chat to moderate problematic viewers), a setting similar to the language setting could be applied as well. If the moderators thought a specific viewer was suspicious, they could be able to click on the viewer's name, and all the messages from this viewer would be highlighted (e.g., in red color) in the following message flow. One click and starting to track the subsequent behavior would amplify their capability of moderating. However, the prerequisite is a setting that can slow down or speed up the chat movement so that moderators can accurately identify the problematic viewers.

Content and rule category setting: this feature is inspired by multi moderation and applied for different channels and content, but it could also apply to any single channel. It means that one bot can have a setting that contains many different rules and streaming content categories that accommodate the norms and guidelines of different channels. From our interviews, the rules for teen channels did not apply to adult channels. In adult channels, adult jokes were permitted but might be inappropriate for teen channels. Hence, settings that can choose a content category first and then apply a rule for that specific category would improve moderation accuracy and avoid embarrassing situations and negative impressions.

LIMITATIONS AND FUTURE WORK

Our sample is only from Twitch. Further research can take other live streaming platforms into account and validate the results. It is also important to note that very few social media platforms have a governance structure in place that allows for third-party moderation tools. That said, it would be interesting to know what kind of moderation tools are being used by companies that do not have third-party tools. Besides, the authors randomly recruited participants on Twitch but finally obtained more male than female and also included some transgender. The biased gender toward male may have

an impact on the part of the results. Since gender difference is beyond the scope of this study, further research may explore the tool or feature preference among these genders. Lastly, many other potential perspectives on the themes of moderation tools can be triggered; For example, future research can explore how to facilitate communication among viewers and mods in the viewer control theme, and chat control theme might need further investigation to understand better how to reduce information overload of mods in the live streaming community.

CONCLUSION

Through the interviews with a diverse sample of moderators on Twitch, the authors used a grounded theory approach and identified four high-level uses of moderation tools that provide a method of conceptual categorization that can potentially apply to any live streaming platforms. Through the summarization of mods' expectation of tools in the future, several functions that can fulfill mods' needs are identified and support the four abstractive perspectives. Since multiparty-based chatbots are underexplored, this research provided many insights into bot development in the live streaming community and raised issues related to social interaction among moderators and viewers, community norm evolution, and technical development of moderation tools. Live streaming is still growing very fast, and content moderation for it is still a challenging issue. No existing bot is perfect to meet the moderator's needs, indicating that there are a potential market and opportunities for related bot development.

ACKNOWLEDGMENT

This project was supported by the National Science Foundation [grant number 1841354].

REFERENCES

Anis, B., Börrnert, F., Rümmeli, M. H., & Kuntscher, C. A. (2013). Role of the pressure transmitting medium on the pressure effects in DWCNTs. *Physica Status Solidi. B, Basic Research*, *250*(12), 2616–2621. doi:10.1002/pssb.201300062

AnyKey. (2016). *Barriers to Inclusion and Retention: The Role of Community Management and Moderation Whitepaper*. Retrieved from http://www.anykey.org/wp-content/uploads/Barriers-to-Inclusion-whitepaper.pdf

Berge, Z. L., & Collins, M. P. (2000). Perceptions of e-moderators about their roles and functions in moderating electronic mailing lists. *Distance Education*, *21*(1), 81–100. doi:10.1080/0158791000210106

Cai, J., & Wohn, D. Y. (2019). Live streaming commerce : Uses and gratifications approach to understanding consumers ' motivations. *Proceedings of the 52nd Hawaii International Conference on System Sciences - HICSS'* (pp. 2548–2557). Retrieved from https://scholarspace.manoa.hawaii.edu/handle/10125/59693

Cai, J., Wohn, D. Y., Mittal, A., & Sureshbabu, D. (2018). Utilitarian and hedonic motivations for live streaming shopping. *Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '18* (pp. 81–88). ACM. doi:10.1145/3210825.3210837

Chadwick, A. (2006). *Internet politics: States, citizens, and new communication technologies*. Oxford University Press. doi:10.1089/153312902753300042

Chancellor, S., Pater, J. A., Clear, T. A., Gilbert, E., & De Choudhury, M. (2016). #thyghgapp: Instagram content moderation and lexical variation in pro-eating disorder communities. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16* (pp. 1199–1211). ACM. doi:10.1145/2818048.2819963

Charmaz, K., & Belgrave, L. L. (2015). Grounded Theory. In The Blackwell Encyclopedia of Sociology. Blackwell. doi:10.1002/9781405165518.wbeosg070.pub2

Clément, M., & Guitton, M. J. (2015). Interacting with bots online: Users' reactions to actions of automated programs in Wikipedia. *Computers in Human Behavior*, *50*, 66–75. doi:10.1016/j.chb.2015.03.078

Coutinho, P., & Jose, R. (2017). Moderation techniques for user-generated content in place-based communication. *Proceedings of the Iberian Conference on Information Systems and Technologies, CISTI*. Academic Press. doi:10.23919/CISTI.2017.7975786

Daniel, L., Bernd, H., & Tom, D. G. (2013). Twitter integration and content moderation in GDACSmobile. *Proceedings of the 10th International ISCRAM Conference* (pp. 67–71). Academic Press. Retrieved from http://publications.jrc.ec.europa.eu/repository/handle/111111111/32413

Delort, J.-Y., Arunasalam, B., & Paris, C. (2011). Automatic moderation of online discussion sites. *International Journal of Electronic Commerce*, *15*(3), 9–30. doi:10.2753/JEC1086-4415150302

Friedländer, M. B. (2017). Streamer motives and user-generated content on social live-streaming services. *Journal of Information Science Theory and Practice*, *55*(11), 65–84. doi:10.1633/JISTaP.2017.5.1.5

Ghosh, A., Kale, S., & Mcafee, P. (2011). Who moderates the moderators? crowdsourcing abuse detection in user-generated content categories. *Proceedings of the 12th ACM Conference on Electronic Commerce* (pp. 167–176). Academic Press. doi:10.1145/1993574.1993599

Gilbert, E. (2013). Widespread underprovision on Reddit. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work - CSCW '13*. Academic Press. doi:10.1145/2441776.2441866

Gilbert, E., & Karahalios, K. (2010). Understanding deja reviewers. *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work - CSCW '10*. ACM. doi:10.1145/1718918.1718961

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85051469782&partnerID=40&md5=8d850b5298b7e5dc1a1fcf4c427fe3da

Hamilton, W. A., Garretson, O., & Kerne, A. (2014). Streaming on twitch: fostering participatory communities of play within live mixed media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1315–1324). Academic Press. doi:10.1145/2556288.2557048

Hammer, H. L. (2017). Automatic detection of hateful comments in online discussion. *Lecture Notes of the Institute for Computer Sciences. Social-Informatics and Telecommunications Engineering, LNICST*, *188*, 164–173. doi:10.1007/978-3-319-52569-3_15

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. doi:10.1177/1049732305276687 PMID:16204405

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *The American Political Science Review*, *107*(2), 326–343. doi:10.1017/S0003055413000014

Lampe, C., & Resnick, P. (2004). Slash(dot) and burn: distributed moderation in a large online conversation space. *Proceedings of the 2004 Conference on Human Factors in Computing Systems - CHI '04*, 543–550. doi:10.1145/985692.985761

Lampe, C., Zube, P., Lee, J., Park, C. H., & Johnston, E. (2014). Crowdsourcing civility: A natural experiment examining the effects of distributed moderation in online forums. *Government Information Quarterly*, *31*(2), 317–326. doi:10.1016/j.giq.2013.11.005

Long, K., Vines, J., Sutton, S., Brooker, P., Feltwell, T., & Kirman, B., … Lawson, S. (2017). "Could you define that in bot terms?": Requesting, creating and using bots on Reddit. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17* (pp. 3488–3500). Academic Press. doi:10.1145/3025453.3025830

Lu, Z., Xia, H., Heo, S., & Wigdor, D. (2018). You watch, you give, and you engage: A study of live streaming practices in China. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. Academic Press. doi:10.1145/3173574.3174040

Mironov, S. V., Faizliev, A., Sidorov, S. P., & Gudkov, A. (2016). Stochastic optimal growth model with s-shaped utility function. *CEUR Workshop Proceedings*, *1623*(12), 234–241. doi:10.1002/pssb.201300062

Müller-Birn, C., Dobusch, L., & Herbsleb, J. D. (2013). Work-to-rule: The emergence of algorithmic governance in Wikipedia. *Proceedings of the 6th International Conference on Communities and Technologies(C&T13)* (pp. 80–89). Academic Press. doi:10.1145/2482991.2482999

Pfaffenberger, B. (2011). "A standing wave in the web of our communications": Usenet and the socio-technical construction of cyberspace values. In C. Lueg & D. Fisher (Eds.), *From Usenet to CoWebs: interacting with social information spaces* (pp. 20–43). doi:10.1007/978-1-4471-0057-7_2

Resnick, P. (2002). Beyond bowling together: Sociotechnical capital in the new millennium. *Human-Computer Interaction*, *77*, 247–272.

Roberts, S. T. (2014). *Behind the screen: The hidden digital labor of commercial content moderation.* [Doctoral dissertation]. University of Illinois at Urbana-Champaign. Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/50401/Sarah_Roberts.pdf?sequence=1

Roberts, S. T. (2016). Commercial content moderation: Digital laborers' dirty work. In S. U. Noble & B. Tynes (Eds.), *The Intersectional Internet: Race* (pp. 147–160). Academic Press; doi:10.1007/s13398-014-0173-7.2

Roberts, S. T. (2017). Content moderation. In Encyclopedia of Big Data. Springer. doi:10.1007/978-3-319-32001-4_44-1

Roger, C. (2018). Twitch bringing ads back to Prime will not endanger its market share despite fan backlash. Super Data Research. Retrieved from https://www.superdataresearch.com/twitch-bringing-ads-back-to-prime-will-not-endanger-its-market-share-despite-fan-backlash/

Saúde, M. R., De Medeiros Soares, M., Basoni, H. G., Ciarelli, P. M., & Oliveira, E. (2014). A strategy for automatic moderation of a large data set of users comments. *Proceedings of the 2014 Latin American Computing Conference - CLEI ' 14*. Academic Press. doi:10.1109/CLEI.2014.6965181

Scheibe, K., Fietkiewicz, K. J., & Stock, W. G. (2016). Information behavior on social live streaming services. *Journal of Information Science Theory and Practice*, *4*(2), 6–20. doi:10.1633/JISTaP.2016.4.2.1

Seering, J., Kraut, R., & Dabbish, L. (2017). Shaping pro and anti-social behavior on Twitch through moderation and example-setting. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17* (pp. 111–125). ACM. doi:10.1145/2998181.2998277

Seering, J., Luria, M., Kaufman, G., & Hammer, J. (2019). Beyond dyadic interactions: Considering chatbots as community members. *Proceedings of 2019 CHI Conference on Human Factors in Computing Systems - CHI ' 2019*. Academic Press. doi:10.1145/3290605.3300680

Seering, J., Wang, T., Yoon, J., & Kaufman, G. (2019). Moderator engagement and community development in the age of algorithms. *New Media & Society.*

Tang, J. C., Venolia, G., & Inkpen, K. M. (2016). Meerkat and Periscope: I stream, you stream, apps stream for live streams. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16* (pp. 4770–4780). Academic Press. doi:10.1145/2858036.2858374

Veglis, A. (2014, June). Moderation techniques for social media content. In International Conference on Social Computing and Social Media (pp. 137-148). Cham: Springer. doi:10.1007/978-3-319-07632-4_13

Wohn, D. Y. (2019). Volunteer moderators in Twitch micro communities: How they get involved, the roles they play, and the emotional labor they experience. *Proceedings of 2019 ACM Conference on Human Factors in Computing Systems*. ACM. doi:10.1145/3290605.3300390

Wohn, D. Y., Freeman, G., & McLaughlin, C. (2018). Explaining viewers' emotional, instrumental, and financial support provision for live streamers. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. Academic Press. doi:10.1145/3173574.3174048

Jie Cai is a Ph.D. student in Informatics at New Jersey Institute of Technology. His current research focused on the live streaming community such as live streaming commerce and content moderation.

Donghee Yvette Wohn is an assistant professor at New Jersey Institute of Technology. Her research area is in human-computer interaction (HCI) and computer-mediated communication. She studies the role of algorithms and social interactions in live streaming, esports, gaming, and social media.